



Freedom, Compulsion, and Causation¹

Jenann Ismael
Centre for Time,
University of Sydney
Department of Philosophy
University of Arizona
jtismael@u.arizona.edu
© J. Ismael 2007

PSYCHE 13/1, April 2007

KEYWORDS: causation, freedom, compulsion, causal phenomenology, asymmetry, intervention, reflexivity, projection, Hume, Pearl, Russell, Cartwright

Abstract: The intuitive notion of cause carries with it the idea of compulsion. When we learn that the dynamical laws are deterministic, we give this a causal reading and imagine our actions compelled to occur by conditions laid down at the beginning of the universe. Hume famously argued that this idea of compulsion is borrowed from experience and illegitimately projected onto regularities in the world. Exploiting the interventionist analysis of causal relations, together with an insight about the degeneracy of one's epistemic relations to one's own actions, I defend a Humean position with regard to the idea of causal compulsion. Although I discuss only compulsion, a similar story could be told about the temporal directedness of causation.

1. Introduction

Much of the literature on the purported clash between personal freedom and a scientific world-view in which action falls under the scope of deterministic dynamical laws concerns questions about the kind of control that decision and deliberation exercise over action. But there is a separable cluster of intuitions about the threat that determinism poses to freedom that have more to do with a species of compulsion that people associate with the notion of cause. For most of us, the notion of cause is rooted in the primitive experience of pushing, holding, kicking and yanking. When you throw a ball across a room, you feel the transfer of force from hand to object. It has a direction and carries with it a feeling of compulsion. When we learn that the dynamical laws are deterministic, we give this a causal reading and we transpose the causal notions drawn from the experience of acting onto the relations between natural events. So we suppose, to use Hume's

favoured example, that one billiard ball *impels* the one that it strikes to move after impact, and to move in the particular direction that it does. Then when we learn that the physical laws are deterministic, we imagine our own actions lying in the wake of a compulsive force emanating from the initial conditions of the universe and the idea of compulsion by external forces threatens our sense of freedom. It is psychologically very difficult to hold in your mind simultaneously the idea that decisions are free and that they are *caused* by facts external to you in the same way that a ball is caused to move by the force of your throw. Even if our decisions exercise regulative control over the production of behaviour, so long as our actions are carried out under the compulsion of causes that were present before we came on the scene, it can seem that we cannot be truly free. From the point of view of this worry, control seems almost irrelevant.

I want to confront this worry by asking where this idea of causal compulsion comes from and whether it is legitimate to project it into the relations between natural events. My conclusion will be that it is not.

1. Physics and causes

Philosophers of mind tend to take it for granted that causal relations are part of the mind-independent fabric of the physical world. But in fact, their status has been hotly contested since Russell famously observed that the closest thing to causal relations in physics are time-symmetric dynamical laws relating global time slices of world-history (Russell 1953: 171-196). These bear a distant relationship to the local, asymmetric relations, with their special brand of necessitation, that form the core of the folk notion of cause. The dynamical laws are differential equations that relate the state of a system at one time to its state at others. They are defined, in the first instance, for the universe as a whole.² Laws are deterministic if later states are completely determined by prior states. They don't describe intrinsic relations between particular events, they have no intrinsic direction, and they don't carry with them anything like compulsion. We can use them to predict earlier states from later ones as surely as we can use them to predict later states from earlier ones. And because the Newtonian laws are deterministic in both directions, while it is true that the state of a Newtonian system at one time determines its state at the next, it is also true that its state at the second determines its state at the first. Dynamical laws are so different from the local, directed transference of force that we associate with the concept of cause that Russell was prompted to write in 1913:

The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm whole (Russell 1913: 1).

Russell himself recommends that we replace the notion of cause with that of a dynamical law. But the matter is not so simple, for reasons that were brought out forcefully in an influential response by Nancy Cartwright (1979: 419-437).³ Cartwright agreed with Russell that causal facts are distinct from (and not reducible to) dynamical facts, but argued that dynamical laws could not play the role of causal relations in deliberation. Agents choosing among potential actions need specifically causal information to distinguish effective from ineffective strategies for bringing about ends. So, for example, it might be true as a matter of physical law (because smoking causes bad breath), that there is a strong positive correlation between bad breath and cancer. But it is not true that

preventing bad breath is an effective strategy for preventing cancer because bad breath does not *cause* cancer. And that difference is not one that submits of a purely dynamical analysis. When we're not merely interested in prediction, but deciding what to *do*, it is the specifically causal information that is needed.

In recent years, there has been a great deal of progress elucidating causal concepts and clarifying their relations to dynamical laws. So much so, in fact, that the tide in the philosophy of science has changed from viewing causal facts as potentially reducible to dynamical information to thinking of them as distinct from, but no less real than, dynamical facts. Judea Pearl describes a shift in his own thinking that exemplifies the change:

[I used to think that] causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world (Pearl 2000: xiii-xiv.)

2. Causal information as information about the results of interventions

A large part of the credit for this change is due to the interventionist account of causal relations, of which Pearl's own work forms a central core. The interventionist account analyses causal information using the notion of an intervention. Dynamical equations, as I said, tell us how the state of the universe at one time depends on its state at another. This means that if we start with information about the past in the form of a probability distribution over past events, we can derive probabilities for events to come. The only strict relations of covariation are among global time-slices, but when we look at how the laws transform probability distributions, we find a rich pattern of correlations among local occurrences.⁴ What causal information adds to this pattern of correlations, according to interventionists, is a separation of a parameter's *effects* from information it carries in virtue of being correlated with past causes from its own effects.⁵ The additional information tells us what *would* happen if we could interrupt the orderly dynamical flow and twiddle, or tweak, or change the value of A. It tells us the effects that variation of A would have if it could be lifted out of the dynamical flow, severed from its own past causes and allowed to vary freely. This lifting a parameter out of the dynamical flow and treating it as a free variable, is what is meant by an "intervention". Hence, the interventionist slogan: information about the causal effects of A is information about the results of hypothetical interventions. It tells us how future probabilities are affected by variation in A when A is treated as a free variable. So in the example of bad breath and cancer, even though there may be a strong lawlike correlation between bad breath and cancer, bad breath is not regarded as a *cause* of cancer if improving people's breath would not alter the cancer rate. Smoking, by contrast, is regarded as a cause of cancer because abolishing smoking, it is believed, would significantly lower cancer rates. This is what it *means* to say that bad breath is *not*, but smoking *is*, a cause of cancer.

For any set of dynamical equations of sufficient complexity, we know that there are multiple conflicting causal hypotheses preserving the correlations among events but

disagreeing over the results of hypothetical interventions. Causal structure separates what a parameter *does* from the information it happens to carry about the future in virtue of common causes in the past. Since that is a difference that only shows up when its own links to other variables are severed, and since no variable is ever actually severed from its causes, this subjunctive content can be captured only in counterfactual terms. Think of a newscast in which Greenspan announcing an increase in interest rates and Bush announcing new ‘security’ measures in the Middle East follows a weatherman forecasting rain. Firing the weatherman won’t ward off rain, but firing Greenspan would likely ward off rises in interest rates and ejecting Bush would, at least in the short term, ward off new measures in the Middle East. That difference is not one that can be made out in terms of correlations between reports and the events reported.

There are various ways of *effectively* severing links to the environment when attention is focused on a proscribed system, so that we can treat changes to the values of some variables as interventions. Any variable that can be manipulated and whose own causes are uncorrelated with the variables of interest can be regarded, for such purposes, as “exogenous”, or external to the system, and changes in the value of such variables as interventions on it. This is the essence of scientific experimentation, as Pearl writes:

The scientist carves a piece from the universe and proclaims that piece *in...* The rest of the universe is then considered *out...* This choice of ins and outs creates asymmetry in the way we look at things and it is this asymmetry that permits us to talk about ‘outside intervention’ and hence about causality (Pearl 2000: xiii-xiv).

But the separation is only an artefact of the restriction of attention. If we widen our view to take in the universe as a whole, there are no exogenous variables. And if there are no exogenous variables, there are no interventions. Pearl acknowledges that:

If you include the entire universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated [lose] their distinction...(Pearl 2000: xiii-xiv).

The universe is a closed system, forming a single, seamless pattern of events. From a god’s eye view, the variables that are treated for experimental purposes as exogenous, including the actions of human agents manipulating them, are just part of the pattern of correlation. We can see now why causal structure doesn’t show up in the pure dynamics. We can also see why information about causal structure would be of little interest for an observer located outside the pattern of events. Such an observer is interested simply in forming beliefs about the pattern, and for these purposes, it is correlations that matter. If the correlations between what the weatherman says and what actually happens is quantitatively the same as the correlation between what Greenspan announces and what actually happens, they have exactly the same role epistemic reasoning. And from a detached perspective, this is the only kind of reasoning there is.

3. Interventions as action from a self-centered perspective

Early work on causation analyzed causes roughly as handles whereby human agents could manipulate effects. It was seen as an advance when the reference to human action was replaced with the notion of an intervention, which—although it couldn’t be analyzed in non-*causal* terms—at least could be characterized neutrally and objectively as a certain

kind of causal process. The general sentiment among interventionists was that reduction to dynamical concepts is not a requirement of causal realism, but objectivity is, because causal realism should require that there are causes in the world even if there are no humans. I'm going to make what looks from this perspective like a retrograde maneuver, reintroducing reference to the self-centered perspective of the participant to explain why causal relations don't show up in physics but are nevertheless indispensable for practical deliberation. It's not an accident that the difference between causal and purely dynamical content shows up not in epistemic reasoning, but in the context of choice. This is already an indication of its perspectival nature, for choice is an attitude that one can only take to one's own actions,

We've seen that causal information is information about the results of interventions, that what counts as an intervention depends on drawing boundaries, and that what boundaries we draw (i.e., which variables we treat as endogenous and which we treat as exogenous) depends on our interests. We also saw that if we widen our view to include the universe as a whole so that nothing lies outside the boundaries, the notion of an intervention ceases to have application. But there is a qualification that I'm going to argue is relevant for understanding the status of causal notions. It is correct that from a god's eye perspective outside of space and time, there are no interventions, but from the self-centered perspective of a system whose activity is *part* of the pattern, its own actions necessarily have the status of interventions. And it is in assessment of the results of those actions that causal thought arises. The reason has to do with degeneracy in the epistemic stance when it is applied reflexively. A system that is representing activity that includes its own will unavoidably encounter the degeneracy. Alethic constraints, i.e., constraints on representation imposed by the way the world is, are what guide representational activity. If I making am telling a story to relate a set of events and I want to be truthful, what I say will be guided by what actually happened. The same goes for other forms of representational activity. If I am making a sketch of a building and I want to be accurate, what I draw will be guided by the shape of the building. These constraints on representational activity are empty when applied reflexively, i.e., when what is being represented is the representational act itself. The most familiar examples of this degeneracy are self-representing linguistic performances: "I promise to X", "I declare that Y". Such performances are perfectly good representational acts. They have truth conditions that can fail to obtain; someone else can certainly falsely ascribe a promise to me, and I can misrepresent my own past promises and declarations. But because they provide their *own* truthmakers, they are unconstrained at the time that they are made. They are self-fulfilling.

Anything can get filled in for X or Y here and being told to *tell the truth* is not going to provide guidance about what to fill in. Being told to tell the truth about what someone *else* will say, or what you said at some other time or place, will constrain your answer, and will lead you to times and places outside the representational act for guidance. But when you are representing your *very own thoughts or statements at the time at which you are having or making them*, the way the world is, together with the desire to tell the truth will provide no guide. One way of putting this is that there exist multiple solutions to the question 'what will I say?' that are inconsistent with one another, but any one of which would provide a correct answer. And no evidence in your

possession could sway your answer, since by performing A you *create* evidence for ‘A’ stronger than any evidence you might have had for not A.

Let’s use the word ‘performance’ for a self-fulfilling representational act. Syntactically, any time you can add ‘hereby’ to a proposition (e.g., “I (hereby) declare that __”, “It is (hereby) acknowledged that __”, “I (hereby) decide to __”...), you have a performance. Decisions are performances. How efficacious one’s decisions are is not up to an agent. I can decide to do any manner of things, but the link between the decision to A and A is a matter of how my decisions are connected in the world and reliability will vary. The link between the decision to raise my arm and the raising of the arm is highly reliable. The link between a decision on my part to produce a delicious mushroom risotto and the production of the risotto is less reliable. And the link between a solitary decision on my part to save the developing world and the saving of the developing world is very much less so. Efficacious action depends on having an accurate understanding of how reliable the connection between ones decisions and the events they concern are. But the decisions themselves – the acts of deciding to A - are self-fulfilling, unconstrained by the requirements of accuracy. This is what I mean when I say that alethic constraints are degenerate when applied reflexively.

It is important the degeneracy is perspectival. It is relative both to agent and to time. “JI decided to A at time t” thought by anyone other than JI, or thought by JI at any other time than while engaged in deliberation is constrained in the ordinary way, on pain of falsity, by whether or not JI decides to A at t.⁶ And it is important that the degeneracy is unavoidable for any system that includes its own activity in the field of representation. The desire to tell the truth in general will not guide my answer the question ‘Will I A?’ and the ordinary epistemic procedures for getting information about whether X A’d will not apply. Guidance has to come from elsewhere. “Will I A?” becomes the question “Should I A?” That leads to the question “What would happen if I B’d or C’d instead?” And it is here that causal information is indispensable.

The emptiness, or degeneracy of alethic constraints (i.e., constraints imposed by truth) when applied to one’s own actions opens up the space for deliberation. I believe that it captures the sense in which, from the point of view of the participant in a dynamical process, her own actions have the status of what Ramsey called “an ultimate contingency’ (and, indeed, as he also said, the *only* ultimate contingency). One’s own contributions to history, in the form of decisions, have a special, degenerate status; they are, in the technical sense, free variables, lifted out of the causal order. And the degeneracy of decision is inherited by the actions they control. Any time you have a representational act that represents an event that is, or is probabilistically dependent, on the act itself, you have the degeneracy.⁷ Orders, also, fit this pattern. Velleman gives the example of a doctor that says to a nurse at the end of an examination: “you will now take the patient to the X-ray room”. The event described here (the conveyance of the patient to the X-ray room²²) is probabilistically dependent on the act of description.⁸ Indeed, it’s quite natural to think of the decision to, e.g., raise one’s arm, as an order that the mind gives to the body. Far from being an esoteric special case, much of the time, in the cases that matter most, when we’re reasoning about the future, we’re reasoning about events that depend either directly, or in an attenuated manner, on our decisions. We ignore the symptoms of degeneracy only by a sharp, and ultimately indefensible, division between

deliberation and epistemic reasoning. Traditional epistemology, and its modern Bayesian incarnation is an epistemology built not for participants, but for detached observers. It is an epistemology built for detached observers because it presupposes the independence of what one is reasoning about from the process of reasoning itself. The embedded agent is rarely in that situation. For the embedded agent, epistemic reasoning is very often, in some part, deliberative.

Anscombe, with characteristic astuteness, noticed the anomalousness of decision from an epistemic point of view. She focused on the closely allied notion of intention and put the difficulty sharply in the form of an observation about the difference between predicting that A will X and forming the intention to X. She writes:

The distinction between an expression of intention and a prediction is generally appealed to as something intuitively clear. ‘I am going to be sick’ is usually a prediction; ‘I am going to take a walk’ is usually an expression of intention.... We might attempt to make the distinction out by saying: an expression of intention is a description of something future ... which description [an agent] justifies (if he does justify it) by reasons for acting... not by evidence that it is true (Anscombe 1957: 1-6).

But she didn’t diagnose the difference. She continues: “But having got so far, I can see nowhere else to go along this line, and the topic remains rather mystifying”.⁹

4. The observer vs. the participant

The point that is important for the purpose of understanding both the need for causal thought and the source of the attendant phenomenology—i.e., *what it is like* to be an embedded, embodied participant in the world, not an onlooker, but one of the players on the field—is the degeneracy of epistemic thought with respect to one’s own activity. It is inside the context of deliberation that the distinction between what an event causes and what it forecasts matters, as we rehearse the possibilities for action, treating them as interventions in the shared environment, assessing their effects and weighing them against our ends. It is here, where truth is no longer a guide, that causal thought has its home. That degeneracy is a perspectival effect. God sees only correlations and conjunction. There is no distinction from an external perspective between what an event forecasts and what it causes (no distinction between the relation the weatherman’s reports bear to the weather and Greenspan’s declarations bear to interest rates). The participant, however, who can intervene and observe the results of interventions, is able to ferret out the network of causal relations supporting the correlations. His epistemological advantage over the spectator consists in the fact that where information flows one way between the spectator and the system of interest, it flows two ways between the participant and the system. The spectator receives data, but the participant is both recipient *and* source. The epistemological advantage that the intervener has over the observer is well documented in both Pearl’s book, and elsewhere. It confirms what we know from experience about the advantage a policeman has, for example, if he can interview a suspect rather than reading a prepared statement.¹⁰

There are trivial and more interesting notions of participation. In the trivial sense, a participant is a system that includes its own activity in its representational scope. This is

enough to yield degeneracy in values of certain variables in the field of representation, and hence enough to make the system unavoidably active with respect to the values of those variables. This sort of degeneracy is a purely formal effect, of which one symptom is the semantic ungroundedness that generates the liar paradox and paradoxes of reference.¹¹ It is just a matter of widening the scope of representation so that the representational relation has fixed points. We get something interesting from a dynamical perspective only when those variables interact dynamically with variables whose values we get information about perceptually. In that case we have the epistemological situation that has the structure we need for causal relations; i.e., not just free variables, but perceptual feedback from the results of tweaking those variables. The crucial thing here is a two-way exchange of information. That two-way exchange is both what gives the participant his epistemological advantage over the observer, and—I want to suggest now—generates a sense of effective presence.

Suppose you are in a video arcade with your hand resting on the console of one of the games while you watch your child dump coins into machines, shoot guns and turn wheels. Something catches your eye and you realize that the restless movements you've been making with your hand are controlling some of the action on the attached screen. It dawns on you that there is unused playing time in the game and that you are actually playing. You discover pedals and with some experimental pushing and pressing, you figure out what the knobs and pedals control, i.e., what changes they effect on the screen, and suddenly you're in the world of the game. Once you've determined how your interventions are connected in this universe, you can begin to participate. You experience yourself as having an effective presence on the screen, directing its course within the parameters provided by the game. In some games, you might do little more than steer a shape around a maze. In others, you might take the action from one locale into another, blow things up, create sounds, and so on. My final suggestion is that this richly structured experience of self as occupant and source of some of the changes that one observes depends on perceptual feedback from the results of interventions. It is the product of a history of interaction that belongs to the participant and not to the spectator, in which feedback between intervention and observation has established robust psychological links between action and observation. It seems to happen largely unconsciously and is mediated at first by phenomenological awareness of presence in the non-virtual world (e.g., by the feel of our hands on the console). But the mind is very quick to make reliable mediating processes phenomenologically transparent. Normally, this sense of effective presence extends to the boundaries of the body, but we know that it can grow with the right feedback to include artificial prostheses and tools (as when a blind person seems to literally feel the ground with the end of his cane), and that it can contract in the absence of that feedback (as when a subject loses voluntary control of limbs).

If this is correct, it is the tight dynamical loop in which the agent is acting an observing the effects of its actions, i.e., the repeatable cycle of willing and watching, intervening and observing the results of intervention) that divides the events one observes into the things one *does* and the things that merely *happen* when one acts, lending a feeling of compulsion to the former. I don't just think that my arm moves *when* I will. It has to move; I *make* it move. It moves *because* I will, in a sense that supports the belief that it wouldn't have moved if I didn't. And I know that because I've tried multiple times, under all kinds of conditions. And I don't just believe that my car accelerates *when* I

press the pedal. It has to accelerate; my pressing the pedal *makes* it accelerate. It accelerates *because* I press the pedal, in a sense that supports the belief that it wouldn't have moved if I didn't. And I know that because I've tried multiple times, under all kinds of conditions. In psychological terms, we know that the feedback is essential to the phenomenology of compulsion. The cycle of intervention and observation generates connections between ideas that are experienced as local, directed, compulsive forces. And we know that the phenomenology can be manipulated by suppressing or altering the feedback, and so we know that it has a psychological source. What all of this suggests a propos of freedom is that there is really no good reason to believe in pushes and pulls built into the fabric of the universe. The pushes and pulls that we experience in acting on the environment relate *ideas* not events. They arise from the flow of information through the self-representational loop from sensation to idea through decision and back out. To this extent, I believe, Hume was right.

5. Conclusion

Summing up, I've suggested that the self-centered perspective from which a system's own actions appear as interventions in the activity it is representing forms the representational context within which the concept of cause arises. It is in this context—the context of *choice* or *decision*—that causal thought, and causal phenomenology has its home. This gives us an account that reconciles Russell's observation that physics doesn't contain causes with Cartwright's insistence on the indispensability of causal information in the context of choice. And, importantly, it discourages the causal reading of the dynamical laws that gives us the idea that if determinism is true, we act under compulsion of a causal force emanating from the initial conditions of the universe.

References

- Albert, D. (2000). *Time and Chance*, Cambridge: Harvard University Press.
- Anscombe, E. (1957). *Intention*, London: Basil Blackwell.
- Beal, J.C. (2003). *Liars and Heaps*, New York: Oxford University Press.
- Cartwright, N. (1979). "Causal Laws and Effective Strategies", *Nous*, 13, p. 419-437.
- Clark, A. (2003). *Natural Born Cyborgs*, New York: Oxford University Press.
- Price, H. and Richard Corry, eds. (2007). *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, Oxford: Oxford University Press.
- Price, H. (2007). "Causal Perspectivalism" In Huw Price and Richard Corry, eds.
- Price, H. (1988). *Facts and the Function of Truth*, Oxford: Blackwell.
- Price, H. (2007). *Naturalism Without Mirrors*, New York, Oxford University Press.
- Ramsey, F (1931), "General Propositions and Causality", in *Foundations of Mathematics and Other Logical Essays*, London: Routledge and Kegan Paul.
- Russell, B. (1953). "On the Notion of Cause", in *Mysticism and Logic*, London: Doubleday.

Velleman, D. (1992). "What happens when someone acts?" *Mind*, 101.

Williams, D. (1950). "Causation and Compulsion", *Journal of Philosophy*, 4.

Notes

1. I'm grateful to Huw Price, to an audience at the Caltech Workshop on Causation in 2005, and to an audience at Kyoto University for very helpful discussion.
2. We can use them to predict the behaviour of parts of the universe by summarizing the relevant part of the state of the whole in what are called boundary conditions. But such predictions are always approximate, always defeasible, always relative to idealizations, and always derivative. The only strict dynamical laws related global time slices of the universe as a whole.
3. See Price & Cory (eds.) (2007) for a selection of papers.
4. Relativistic considerations don't make a difference here.
5. A parameter is just a variable feature of a physical system: a family of properties, exactly one of which pertains to the system at any given time.
6. 'Did I decide on the blue sofa yesterday?' 'Will I decide not to go through with it when the time comes?' are perfectly ordinary questions constrained, respectively, by what I did decide yesterday, and what I will decide tomorrow.
7. The sort of dependency in question is asymmetric probabilistic dependency, conditional on what a system knows about its history, assuming the ordinary epistemological asymmetries between past and future. See Albert (2000), and references there for discussion of those asymmetries.
8. The probability that the nurse takes the patient to the X-ray room is higher, given the doctor's order, than it would have been, we can suppose, had the doctor said "you will now take the patient back to his room."
9. Velleman (1992), picking up Anscombe's discussion, has characterized an intention, in a way that correctly identifies the reflexive source of degeneracy, as a self-referring mental act that causes what it refers to.
10. The advantage is considerable when 'no comment' isn't an allowed answer, there is no telling of lies, and the respondent doesn't get to choose the questions. That is the kind of control that an unconstrained ability to intervene would give us.
11. For discussion of semantic ungroundedness see Beal (2003).